AD_____

Award Number: DAMD17-00-1-0318

TITLE: Markers of Increased Risk in Pre-Invasive Breast Cancer

PRINCIPAL INVESTIGATOR: Soma Mandal, Ph.D.
Peter Watson, Ph.D.

CONTRACTING ORGANIZATION: University of Manitoba
Winnipeg, Manitoba R3E 0W3
Canada

REPORT DATE: February 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 2004	Annual Summary (1 Jan 2003 - 1 Jan 2004)

4. TITLE AND SUBTITLE
Markers of Increased Risk in Pre-Invasive Breast Cancer

5. FUNDING NUMBERS
DAMD17-00-1-0318

6. AUTHOR(S)
Soma Mandal, Ph.D.
Peter Watson, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Manitoba
Winnipeg, Manitoba R3E 0W3
Canada

E-Mail: mandals0@cc.UManitoba.CA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

The understanding of the risk of recurrence and progression of breast cancer, particularly, from a pre-invasive lesion is evasive and challenging. Our goal has been to identify changes in early breast cancer using techniques to profile global gene expression. In previous work, we have employed cDNA array techniques to identify relatively small changes in gene expression between low and high risk pre-invasive cancer. We have therefore optimized the alternative SAGE (serial analysis of gene expression) technique to also apply to this problem and compare with our cDNA results. We have tested SAGE initially in a small pilot experiment, using transfected cell lines that differ in their expression of the psoriasin gene, that is commonly overexpressed in pre-invasive cancer. A pilot experiment using two SAGE libraries showed that differences can be detected and larger experiment is now underway. Clones with inserts have been screened by PCR and agarose gel analysis and are now being sequenced. We anticipate that this experiment will establish the capability of this assay to detect differential gene expression and lay the foundation for its application to pre-invasive tumor samples, and may identify gene that are overexpressed in pre-invasive disease.

14. SUBJECT TERMS
Breast cancer, psoriasin, SAGE, Global gene expression profile

15. NUMBER OF PAGES
20

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

A series of cumulative genetic changes resulting in the suppression of tumor suppressors and overexpression of tumor promoters leads to the transition from the pre-invasive to the invasive stage of a tumor. In recent years there has been an increase in the numbers of ductal carcinoma *in situ* (DCIS) and other pre-invasive lesions diagnosed and it is a significant problem to manage such patients with pre-invasive lesions. Thus, the understanding of genetic interactions occurring in the pre-invasive stage is crucial. It is challenging and necessary to identify molecular markers that can predict the risk of recurrence and/or the progression of pre-invasive to invasive tumors.

Recognizing the fact that genes associated with low and high risk of DCIS need a great accuracy of prediction, we have applied the microarray technique to profile gene expression in a set of DCIS cases. However, we realized the need for a greater accuracy and a global view of the overall gene expression pattern, which we found to be absent in the array analysis. Before applying another technique to the DCIS samples, from which it is difficult to obtain reasonable amounts of RNA, we decided to optimize another technique, SAGE or serial analysis of gene expression, in a breast cancer model cell line. So, we used a wild-type (MDA-MB-231) and its transfected counterpart that overexpress a gene, psoriasin. We reasoned that if SAGE could be effectively used in our hands, we could very easily apply it to profile DCIS cases. We chose the psoriasin gene because of its documented importance in pre-invasive breast cancer lesions.

Psoriasin (S100A7), a calcium binding protein of the S100 family, is a differentially expressed gene in breast cancer, which is highly expressed in pre-invasive ductal carcinoma *in situ* (DCIS), and is persistently expressed in a subset of poor prognosis invasive carcinomas, but is low or undetectable in normal breast and in benign lesions (1, 2, 3). In general, conditions such as growth factor deprivation, hyperproliferation, high lymphocytic infiltration, and cell detachment from the extracellular matrix, all trigger psoriasin expression in DCIS and psoriasis (4, 5). Psoriasin has been shown to up-regulate other proteins that enhance breast cancer progression. We have demonstrated that one such candidate is the human p38Jab1/CSN5 (6). Jab1 is expressed in primary invasive breast tumors (our own unpublished results and 7). Psoriasin promotes the nuclear localization of Jab1 and

enhances tumor progression, and has been shown to influence at least some downstream genes that have been described as Jab1 genes in other non-breast and non-epithelial cell systems (6). Thus, creating an expression profile in the cell line model would pave the way for future use of SAGE for DCIS study.

OBJECTIVES

The objectives of this study are to identify genetic markers associated with pre-invasive lesions, such as DCIS. This will enable the prediction of the risk of recurrence and differentiate between the high and low risk psoriasin expression in breast cancer. The specific aims are:

Specific aim 1: To identify and clone genes that are differentially expressed between high and low grade DCIS that may contribute to their known risks of recurrence.

Specific aim 2: To study the role of candidate genes identified in specific aim 1 by assessment of expression in vivo and by manipulations of expression in breast cancer cell lines.

BODY OF REPORT

The overall goals attained in the context of specific aims are defined in this body of the report. Previously we have used arrays to identify differentially expressed genes in the pre-invasive lesions in many breast cancer cases.

During the course of the ongoing work, we felt the need for a more precise identification of the markers associated with breast cancer. In our previous work,

we had chosen DCIS samples from all grades (high, intermediate and low grades respectively) to discriminate the differential gene expression between necrotic and non-necrotic lesions found in these samples. Using cDNA filter arrays to profile gene expression in pre-invasive DCIS breast lesions, we have shown that there were consistent differences in expression in a subset of genes between the high and the low grade DCIS (8). However, filter array technique has some limitations. Comparing microarray experiments from different laboratories is difficult due to a number of random and systematic errors that are difficult to eliminate. Another problem we faced is that the relative fold difference between the gene expression between groups, though consistent, was relatively small, mostly between 1.5-3.0, with a few exceptions, and we had to apply a confidence level of 75%, which was very lax. Added to the dilemma was the fact that only a relatively small and limited number of changes in gene expression were seen between groups (8). Also, microarrays are subject to false positives and gives no indication of genes that cannot be hybridized to a set of genes spotted on the arrays, not to mention of unidentified genes. Nevertheless, we were able to show with certainty changes in a few genes, such as angio-associated migratory cell protein (AAMP) in our DCIS samples. On the other hand, parallel work by Polyak (4) and others (9, 10) have shown that a more quantitative and global approach to gene profiling could be obtained through SAGE analysis to predict more accurately the genes up- or down-regulated in the different stages of breast cancer (5, 11). Moreover, the sequencing aspect of SAGE gives it a unique advantage. Its digital database facilitates direct comparisons between SAGE libraries. The ability to query a number of SAGE

libraries submitted by various laboratories via the internet represents a powerful resource available to everyone in the scientific community (12). The confidence level is very high and directly correlates to the relative abundance of genes. Thus, a quantitation of genes with accuracy is possible. This prompted us to focus our efforts on the optimization of the SAGE technique in our laboratory with the goal of using this as an alternative technique for profiling gene expression. We anticipate that this experiment will establish the capability of this assay to detect differential gene expression and lay the foundation for its application to pre-invasive tumor samples, and may identify genes that are overexpressed in pre-invasive disease. Psoriasin is one such gene, which is highly expressed in pre-invasive lesions of the breast.

SAGE is a powerful technique, which allows for the evaluation of cellular mRNA populations (transcripts) in cell lines or isolated tissues without prior knowledge of gene sequence (13, 14). Techniques such as cDNA subtraction or differential display are quite useful for comparing the gene expression between two cell types, but provides only a partial picture with no direct information of the gene abundance (13). The expression patterns are deduced from the abundance of the individual tag corresponding to a representative gene. The software extracts SAGE tags from raw sequence files, compares tag abundances between libraries, and also calculates statistically significant differences between libraries by Monte Carlo simulation. Individual tags are identified by comparison with the reference sequence database (SAGE map, http://www.ncbi.nlm.gov/).

KEY ACCOMPLISHMENTS

The experimental model and optimization of SAGE:

The principle model that we have used is the human breast cancer cell line MDA-MB-231. The rationale was: (a) this is an estrogen receptor negative cell line and therefore represents the type of invasive tumors that have expression of psoriasin in tumor samples, but unlike an alternative cell line such as MDA-MB-468, it has no endogenous psoriasin expression (by RT-PCR and Western blot). (b) We have demonstrated that an interaction between psoriasin and Jab1 downstream genes. We have previously cloned the full psoriasin protein coding sequence into pcDNA3.1 for transfection and now possess stable transfectants of MDA-MB-231 cells and high levels of psoriasin expression (HP2). (c) These psoriasin expressing MDA-MB-231 cells express comparable levels of psoriasin as found in cell lines with endogenous expression (MDA-MB-468) and in breast tumors (when compared by RT-PCR and Western blot).

Total RNA was extracted using Triazol reagent (Invitrogen) from wild-type MDA-MB-231 and psoriasin expressing clone, HP2, cultured and harvested in parallel under identical growth conditions. The starting material for SAGE is RNA, which was isolated and quality checked by (a) determination of the 260/280 ratio, and (b) by running the extracted RNA in formaldehyde containing denaturing gels.

The SAGE technique is very useful for the direct quantitation of genes expressed in a given tissue or cells. However, the technique is complex and involves a number of steps, each of which are challenging. Technically, the optimization process was laborious and time-consuming as we ventured to use it in

our facility, where it was not done before. Thus, optimization involved a total set-up and pooling of the resources in our laboratory, as well as troubleshooting of the various stages of the technique.

SAGE was started by capturing the poly (A) mRNA to the oligo(dT) magnetic beads and cDNA was synthesized directly from the magnetic beads. The advantage of using these beads is the prevention of sample loss during the subsequent washing steps. The steps from cDNA synthesis to the cleaving of cDNA by the anchoring enzyme Nla III to linker ligation, and tag synthesis by the tagging enzyme BsmF I, were essentially a single tube procedure, and had to be performed with extreme caution to minimize and prevent loss of beads bound to the cleaved DNA. The synthesized tags were ligated to form ditags with attached 40 bp linkers or adaptors, which were PCR amplified with primers designed in accordance with compatible sequences in the linkers. Since our ultimate goal is to use SAGE for profiling DCIS samples, we are aware of the reality that the amount of available RNA would be limited, to the extent as it is required for a successful and accurate gene quantitation (2.5-5 μg poly(A)$^+$ RNA). So, we have tried, for future experiments, the approach with modifications proposed by Datson *et al* (15). So, after the initial PCR amplification of the ditags, we optimized further the number of cycles for re-PCR between 10-18 to maximize the yield of cDNA without introducing amplification bias. There are other approaches too, which we have yet to try to maximize yield of cDNA. One of which is RNA amplification, but this has been mostly used in array experiments, so it will be a trial and error process. Of course, since we had sufficient amounts of RNA material available from cell lines,

we constructed libraries with straight PCR amplification to generate enough ditags for our pilot experiment.

From this point onwards, sequential gel purification steps were performed leading from ditag isolation to concatemer formation and isolation respectively. To obtain a greater resolution of ditags, we used higher percentages of polyacrylamide gels, however, we found that greater than 12% gels did not improve the resolution significantly. So, we ran PCR amplified ditags on a 12% polyacylamide gel, the 100 bp band comprising of the ditag + linker was cut out from the gel and purified by ethanol precipitation. Isolation without contamination of the 100 bp ditag from a very close 80 bp band (comprising of linkers) was very critical for the future enzymatic steps and had to be done with repeated practice exercising caution.

The linkers were cleaved away from the ditags by Nla III digestion in a 12% gel to yield a 26-28 bp ditag. The purity of the ditags without the linkers was crucial because contamination would potentially poison the ligation reaction in the final step. The optimal time of gel run and additional purification steps were applied to ensure that the ditags were free from the linkers. Additional caution was also exercised to ligate the ditags with promptness. The 26 bp was excised carefully from the gel to separate from the products of partial digestion with promptness; ditags are unstable in the presence of low salt and have to be preserved with high salt (TE was used instead of LoTE for resuspension of the purified ditags) and were always kept below or at 4°C during the purification process. The ditags were purified by ethanol precipitation and the freed ditags were ligated to form concatemers and run on an 8% polyacrylamide gel. The bands were cut out in

various ranges, between 300-1000 bp sizes and purified by ethanol precipitation (steps to the formation of concatemers, Appendix Ai). This concatenated DNA contains tags serially ligated to each other, which were ready to be cloned into a suitable vector. The vector used conventionally for cloning SAGE inserts is pZero-1 (16). The concatenated DNA was cloned into the linearized vector and was propagated in the electrocompetent TOP10 bacterial strain and screened for colonies with inserts by PCR. Initially we had difficulties in cloning inserts with sizes greater than 300-400 bp. We were cautious in overdigesting the vector with Sph I due to the exonuclease activity of the enzyme, until we could optimize the digestion time to yield a greater proportion of the digested vector with overhangs available to take up the SAGE inserts with the compatible overhang. We have now been able to successfully clone inserts with bigger sizes (800-1000 bp). Greater the size of the insert, greater is the possibility of more tags being present in the insert-containing clone. Most of the clones obtained were positive and ranged from 0.5-1.0 kb (Appendix Aii). This way a test library from the wild-type MDA-MB-231 cells was prepared and a few clones were sequenced to verify the optimization of the technique in our hands.

Pilot experiment-construction and analysis of SAGE libraries:

Two SAGE libraries were constructed from wild-type MDA-MB-231 cells and MDA cells transfected with psoriasin (HP2). Each library consisted of ~500 clones. After a detailed screening process, a limited number of positive clones were

purified and samples were submitted for sequencing to the National Research Council in Halifax, Nova Scotia, Canada.

The sequence data files from the SAGE libraries are processed for analysis by using the SAGE software, SAGE 2000, the SAGE extraction software. The primary data product of the SAGE technique is the clone insert sequence, which represents the concatenated tags in pairs, i. e. the ditags. The ditags are separated by 4-base pair punctuation marks, CATG, which are the Nla III recognition sites. The software locates these sites within the ditag concatemer and defines the 10 bp tags from each ditag falling between these sites. The product of this processing is a list of tags with their corresponding tag values, and thus is a digital representation of the cellular gene expression. Together, the final analysis involves a combination of 3 softwares, namely SAGE 2000, Microsoft Access and Microsoft Excel respectively. For comparison, the raw sequence data from the SAGE map public database can be downloaded from the NCBI site.

Analysis of the initial limited samples in the two libraries yielded more than 150 tags per library on an average. Precisely, there were 160 tags for wild-type MDA-MB-231, and 237 tags in HP2. ~30% of the genes were up-regulated in HP2 and ~15% of the genes were down-regulated in HP2 and the rest remained unchanged. Among the up-regulated genes, more than 65% of the genes were present only in HP2, but completely absent in wild-type MDA-MB-231. The number of tags showing increase was more in the HP2 cells than in the wild-type (Appendix B). More samples of positive clones are now being sent for sequencing to obtain a

profile of the differences in the gene expression resulting from overexpression of psoriasin.

Some of the genes showing highest increases in HP2 are:

GAPDH,

triose phosphate isomerase I,

insulin-like growth factor receptor-1,

enoplasmic reticulum protein retention receptor I,

some ribosomal proteins.

In general, ribosomal proteins are absent or some are expressed at low levels in the wild-type MDA cells. Some of the genes expressed in the wild-type MDA line, but not in HP2 are:

eukaryotic translation elongation factor 1 alpha,

translationally-controlled tumor protein 1,

microsomal glutathione S-transferase 1,

histone deacetylase 2,

inositol(myo)-1(or 4)-monophosphatase 1,

S-adenosylhomocysteine hydrolase,

ornithine decarboxylase antizyme.

CONCLUSIONS

Our pilot experiment demonstrates that we can successfully complete the SAGE procedure to identify gene expression from tags from small RNA samples.

Initial trends after analysis of the two libraries point to increases in genes in HP2, which may potentially contribute to the enhanced proliferative potential due to psoriasin expression. However, in our very small experiment we did not identify psoriasin as differentially expressed. This may be a reflection of the overall expression of the transgene relative to other genes, or the very small size of the number of tags in the initial dataset. Given the complexity of the human genome and the libraries generated from it, observing a given gene qualifies as a rare event, as the abundance of most individual message is the order of a few percent or less. Therefore, increasing the number of clones, and hence the number of tags, will increase the chances of detection with more stringent statistical accuracy. Also, we will be able to apply the 11^{th} and the 12^{th} base analysis of tags that may be carrying the same sequences for different genes in the tag-to-gene assignments to determine unambiguously the correct gene, which includes psoriasin as well. Again, this will only be possible with an increase in the size of clones that are being sequenced and which, at the present moment we are engaged in. With data on the differential expression of genes resulting from two cell lines in hand, we plan to successfully apply SAGE to profile genes altered in pre-invasive DCIS breast tumor samples.

REFERENCES:

1. Moog-Lutz, C., Bouillet, P., Regnier, C. H., Tomasetto, C., Mattei, M-G., Chenard, M-P., Anglard, P., Rio, M-C., and Basset, P. (1995). Comparative expression of the *psoriasin (S100A7)* and *S100C* genes in breast carcinoma and co-localization to human chromosome 1q21-q22. Int. J. Cancer, **63**, 297-303,

2. Leygue, E., Hiller, T., Snell, L., Hole, K., Dotzlaw, L., Watson, L., and Watson, P. H. (1996). Differential expression of *psoriasin* mRNA between *in-situ* and invasive human breast cancer. Cancer Res., **56**, 4606-4609,

3. Emberley, E. D., Niu, Y., Njue, C., Kliewer, E. V., Murphy, L. C., and Watson, P. H. (2003). Psoriasin (S100A7) expression is associated with poor outcome in estrogen receptor-negative invasive breast cancer. Clin. Cancer Res. **9**, 2627-2631.

4. Enerbäck, C., Porter, D. A., Seth, P., Sgroi, D., Gaudet, J., Weremowlcz, S., Morton, C. C., Schnitt, S., Pitts, R. L., Stampl, J., Barnhart, K., and Polyak, K. (2002). Psoriasin expression in mammary epithelial cells *in vitro* and *in vivo*. Cancer Res., **62**, 43-47.

5. Porter, D., Lahti-Domenici, J., Keshaviah, Y., Bae, Y. K., Argani, P., Marks, J., Richardson, A., Cooper, A., Strausberg, R., Riggins, G. J., Schnitt, S., Gabrielson, E., Gelman, R., and Polyak, K. (2003). Molecular markers in ductal carcinoma *in*

situ of the breast. Mol. Cancer Res. **1**, 362-375.

6. Emberley, E. D., Niu, Y., Leygue, E., Tomes, L., Geitz, D., Murphy, L. C., and Watson, P. H. (2003). Psoriasin interacts with Jab1 and influences breast cancer progression. Cancer Res. **63**, 1954-1961.

7. Kouvaraki, M. A., Rassidakis, G. Z., Tian, L., Kumar, R., Kittas, C., and Claret, F-X. (2003). Jun activation domain-binding protein 1 expression in breast cancer inversely correlates with the cell cycle inhibitor p27 ^{KIP1}. Cancer Res. **63**, 2977-2981.

8. Adenyinka, A., Emberley, E. D., Niu, Y., Snell, L., Murphy, L. C., Sowter, H., Wykoff, C. C., Harris, A. L., and Watson, P. H. (2002). Analysis of gene expression in ductal carcinoma in situ of the breast. Clin. Cancer Res. **8**, 3788-3795.

9. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. (1997). Gene expression profiles in normal and cancer cells. Science. **276**, 1268-1272.

10. Hibi, K., Liu, Q., Beaudry, G. A., Madden, S. L., Westra, W. H., Wehage, S. L., Yang, S. C., Heitmiller, R. F., Bertelsen, A. H., Sidransky, D. and Jen, J. (1998). Serial analysis of gene expression in non-small cell lung cancer. Cancer Res. **58**, 5690-5694.

11. Porter, D., Krop, I. E., Nasser, S., Sgrroi, D., Kaelin, C. M., Marks, J. R., Riggins, G., and Polyak, K. (2001). A SAGE (serial analysis of gene expression) view of breast tumor progression. Cancer Res. **61**, 5697-5702.

12. Patino, W. D., Mian, O. Y., and Hwang, P. M. (2002). Serial analysis of gene expression. Technical considerations and applications to cardiovascular biology. Circ. Res. **91**, 565-569.

13. Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W. (1995). Serial analysis of gene expression. Science. **270**, 484-487.

14. Velculescu, V. E., Vogelstein, B., and Kinzler, K. W. (2000). Analyzing unchatered transcriptomes with SAGE. Trends in Genetics. **16**, 423-425.

15. Datson, N. A., Jong, J. van derP-de, Berg, M. P. van den., Kloet, E. R. de, and Vreugdenhil (1999). MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. Nucleic Acids Res. **27**, 1300-1307.

16. Angelastro, J. M., Ryu, E. J., Törõcsik, B., Fiske, B. K., and Greene (2002). Blue-white selection step enhances the yield of SAGE concatemers. BioTechniques. **32**, 484-486.

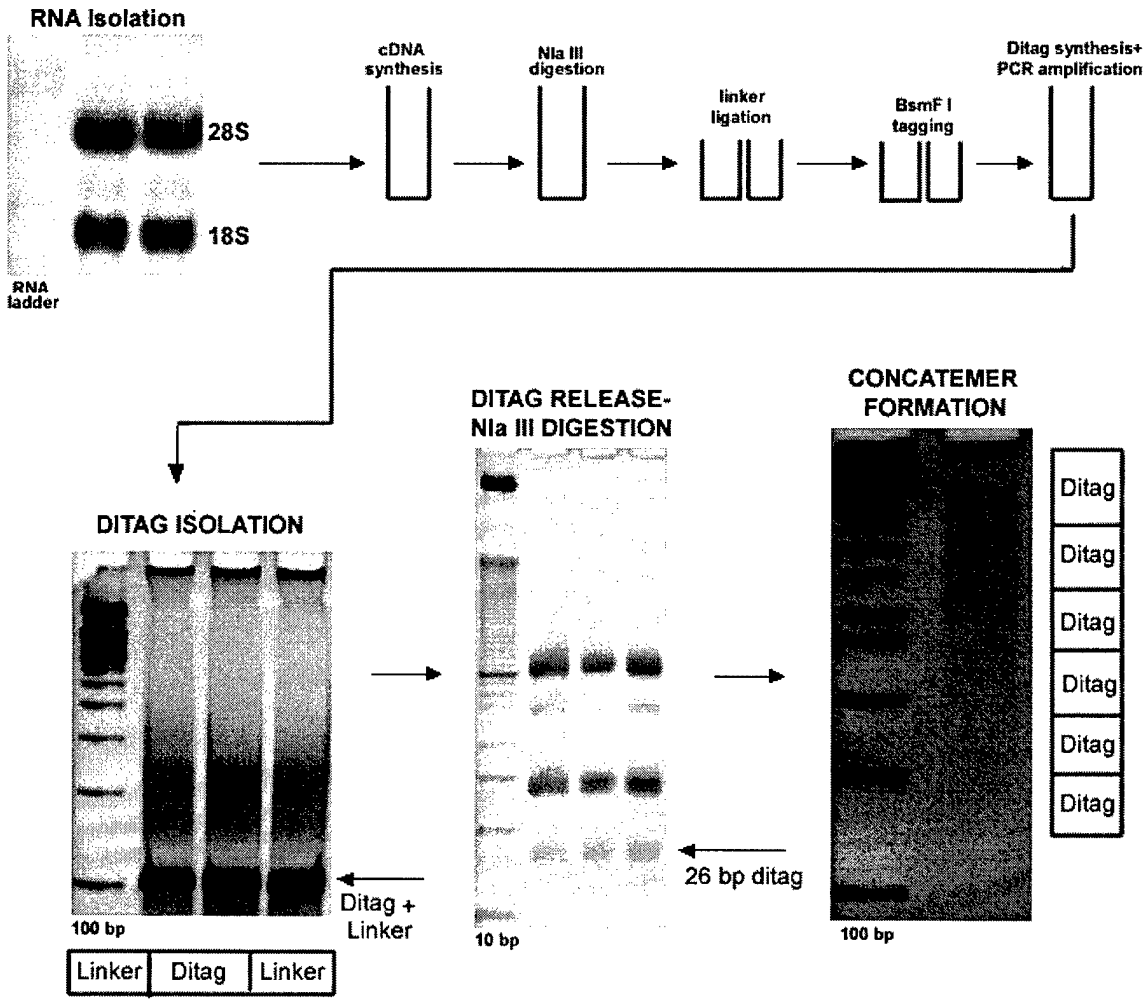
Legends:

Appendix Ai: Steps to the formation of concatemers. RNA was isolated from the cell lines and subjected to denaturing gel electrophoresis to check the integrity of the prep. Next all steps beginning from cDNA synthesis to the formation of tags by a combination of Nla III and BsmF I synthesis were carried out in a single tube and was followed by linker ligation and ditag synthesis and PCR amplification. Amplified ditags were purified in a 12% polyacrylamide gel and later separated from the 40 bp linker in a 12% polyacrylamide gel. The 26 bp ditags were serially ligated and run on an 8% gel and purified.

Appendix Aii: The gel-purified ditags isolated and purified according to size, were cloned into the pZero-1 vector and screened by PCR for positive clones. PCR products were run in a 1.5% agarose gel for the screening of positive clones with various sizes.

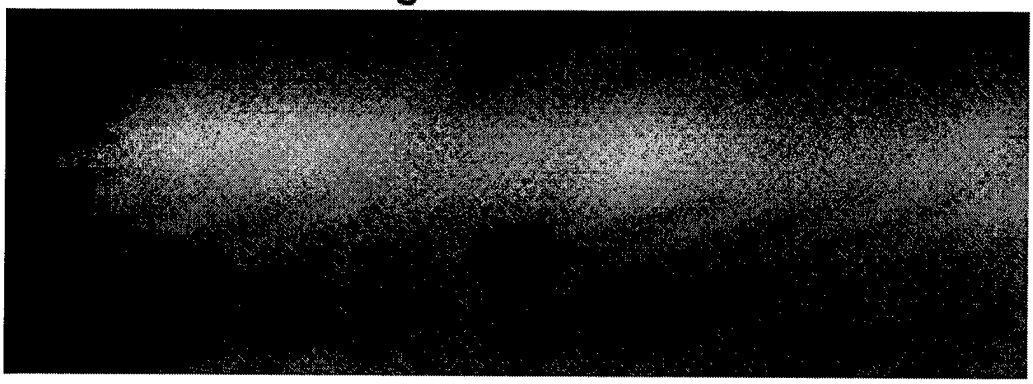
Appendix B: Graphical representation of the percentage abundance of tags in the two SAGE libraries. In general, more tags showed greater abundance in the HP2 cells than the wild-type MDA-MB-231 cells.

Appendix Ai: Steps to the formation of concatemers.



Appendix Aii

Screening Clones for Inserts



Appendix B: Tag abundance in a sample of clones from the HP2 and MDA-MB-231 libraries